DOCUMENT RESUME

ED 124 590 | TM 005 348

AUTHOR	Wonvey, John J.
TITLE	Determining School Effectiveness Following a Regression Analysis.
PUB DATE	[Apr 76]
NOTE	22p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Comparative Analysis; Cost Effectiveness; *Mathematical Models; *Measurement Techniques; Multiple Regression Analysis; Prediction; *Program Effectiveness; *Schools; *Statistical Analysis
IDENTIFIERS	Hyperbolic Confidence Bands; Linear Confidence Bands; Performance Indicators

ABSTRACT

Three methods that can be used subsequent to a regression analysis to determine the relative effectiveness of schools are Dyer's Performance Indices, Scheffe's hyperbolic confidence bands, and Gafarian's linear confidence bands. These methods were applied to data from 54 hypothetical schools randomly generated from a multivariate normal distribution using parameters from previous studies. Data points having Performance indices of 1 and 5 generally fell outside of the Scheffe confidence bands. The linear confidence bands were much wider than the Scheffe bands near the mean and slightly narrower at the extremes. Overall, Performance Indices and Scheffe bands produced similar results. (Author)

ED124590

11.03

Determining School Effectiveness

Following A Regression Analysis

John J. Convey

Catholic University of America

TM005 348

Presented at the Annual Meeting of the

American Educational Research Association,

San Francisco, April, 1976

Determining School Effectiveness

Following A Regression Analysis

Abstract

Three methods that can be used subsequent to a regression analysis to determine the relative effectiveness of schools are Dyer's Performance Indicators, Scheffé's hyperbolic confidence bands, and Gafarian's linear confidence bands. These were applied to data from 54 hypothetical schools randomly generated from a multivariate normal distribution using parameters from previous studies. Data points having Performance Indicators of 1 and 5 generally fell outside of the Scheffé confidence bands. The linear confidence bands were much wider than the Scheffé bands near the mean and slightly narrower at the extremes. Overall, Performance Indicators and Scheffé bands produced similar results.

Determining School Effectiveness
Following A Regression Analysis

Several statistical models have been used in attempts to determine the relative effectiveness of schools. Marco (1974) examined five such models which use longitudinal data. Four of the models involved a regression procedure; the other involved a comparison of mean difference scores. Some evidence indicates that a simple regression model using an initial achievement mean as predictor and a subsequent achievement mean as criterion produces an adequate measure of relative effectiveness in a cost-effective sense (Convey, 1975). This basic model frequently has been employed in field studies (Burke, 1972; Dyer, Linn, & Patton, 1969; Maryland State Department of Education, 1975).

Once a basic regression analysis is completed, the question remains as to what subsequent analyses would be needed to classify adequately the relative effectiveness of the schools involved. The purpose of this paper is to examine three methods that can be used subsequent to a regression analysis, and to determine whether one is best in a cost-effective sense. The methods are: 1) Performance Indicators suggested by Dyer, Linn, & Patton (1967) in a

feasibility study for the New York State Assessment Program; 2) hyperbolic confidence bands about the regression line (Scheffé, 1959); and 3) linear confidence bands about the regression line defined on the subset of interest (Gafarian, 1964). Simulated data were used so that school parameters could be manipulated in order to make some schools more effective than others according to an established criterion.

Performance Indicators

Dyer, Linn, and Patton (1967) in extending a concept introduced by Dyer (1966) generated a general methodology for the calculation of Performance Indicators (PIs). In a given group of schools, the regression of final performance on initial performance and other relevant variables is obtained. Residuals are obtained and an index (I) is computed as follows:

$$I = \frac{\text{residual}}{\frac{\overline{SD}}{(\bar{n})^{\frac{1}{2}}}} \qquad (1)$$

where, $\overline{SD}$ is the average within-school standard deviation on the final performance measure, and $\bar{n}$ is the average number of students per school. PIs are defined as follows:

$$\begin{aligned} I &< -1.5, \ PI = 1; \\ -1.5 \le I &< -.5, \ PI = 2; \\ -.5 \le I &< .5, \ PI = 3; \\ .5 \le I &\le 1.5, \ PI = 4; \\ 1.5 < I &, \ PI = 5. \end{aligned} \qquad (2)$$

The PIs are used then to identify schools that seem to be performing either above expectation or below expectation with respect to a particular class of educational outcomes.

Hyperbolic Confidence Bands

Confidence-banding a regression surface requires the construction of two functions, based on the sample data, which lie entirely above and below the unknown true regression surface with a specified probability. The general problem was solved by Scheffé (1959). An excellent discussion of the procedure appears in Miller (1966). For a single predictor, the familiar Working-Hotelling band (Working & Hotelling, 1929) is a special case of the general Scheffé procedure. The band is given by:

$$\pm(2F_{2,n-2})^{\frac{1}{2}}\, s\left(\frac{1}{n} + \frac{(X - \bar{X})^2}{SS_X}\right)^{\frac{1}{2}} \qquad (3)$$

where, s is the standard error of estimate, n is the number of observations, SS_X is the predictor sum of squares, X is a predictor value, $\bar{X}$ is the predictor mean, and F is the critical value at the appropriate confidence.

The confidence band given by Equation 3 will consist of hyperbolic curves about the regression line with the minimum value occurring at the mean of the predictor. In the remainder of this paper, these bands will be referred to as Scheffé bands.

Linear Confidence Bands

Gafarian (1964) showed how to construct a confidence band of uniform width over a finite interval for a one-predictor linear regression model. Gafarian provides tables of critical values necessary to implement this technique for situations in which the sample size is even and the mean of the predictor lies at the midpoint of the interval over which the confidence band is to be constructed. The resulting band is uniformly wide over the interval of interest.

Given the tables provided, the procedure is straightforward. The table is entered for a particular value of sample size, confidence level, and $c = 2SS_x/n(b-a)$, where a and b are the endpoints of the interval of interest. For $c \geq 1$, the table yields a value of $d(n)^{\frac{1}{2}}$; for $c < 1$, a value of $cd(n)^{\frac{1}{2}}$. The interval is given by $\pm$ ds, where s is the estimated standard error of estimate.

Gafarian shows that these bands generally will be wider than the Working-Hotelling bands around the middle of the interval and narrower at the extremes. Miller (1966) indicates that the Gafarian technique seems to be better for short intervals, and the Working-Hotelling technique better for longer intervals.

Sample

Longitudinal data for 9087 individuals in 54 hypothetical schools were randomly generated from a multivariate normal distribution. The input and output variables were given the characteristics of the total math score in the sixth and eighth grades, respectively, on the Comprehensive Tests of Basic Skills, Level 3, Form Q and Form R, respectively (California Test Bureau/McGraw-Hill, 1970). The expanded standard score scale provided by the publisher was used. The data-generating procedure is described below in detail.

Insert Table 1 about here

First, 54 ordered sets representing scores for input and output were generated according to the specifications given in Table 1. There is considerable empirical evidence to indicate that, for achievement tests, the standard deviation of the distribution of school means is from .3 to .6 of the standard deviation of the distribution of individual scores, regardless of school size (Lindquist, 1930; Lord, 1959). The approximation of .4 suggested by Lord (1959) was used in the specifications. The correlation between input and output is consistent with the findings of Dyer, et al. (1969).

Second, the 54 input scores were ordered from high to low. The highest 18 scores were designated as high, the next

18 as medium, and the lowest 18 as low. Within each category, six groups randomly were designated to be effective groups, six to be average groups, and six to be less effective groups. Effectiveness was defined in terms of the gain from the input mean to the output mean. The output means were paired with the input means so as to satisfy the effectiveness criteria of effective (gain greater than 68), average (gain between 46 and 68), and less effective (gain less than 46). In this study, 68 represented approximately one standard deviation on the input distribution for the individual scores. These criteria appear to be consistent with previous studies (see, Coleman, et al., 1966; Guthrie, 1970; Shaycoft, 1967). In addition, care was exercised to maintain the correlation between the variables approximately at .73. Table 2 shows the characteristics of the ordered sets after pairing.

Insert Table 2 about here

The input-output pairings were made so that within each effectiveness classification, groups with high, medium, and low inputs were represented equally. Use of this procedure attempted to control for any bias that might be introduced by an overbalance of certain levels of inputs in any one classification.

The next step in the data-generating procedure was to establish individual data within each group. Prior to generating the individual scores, group size was varied according to a plan providing 18 groups of 20 to 99, 18 groups of 100 to 199, 9 groups of 200 to 299, and 9 groups of 300 to 399. This distribution is consistent with field results (Florida Ninth-Grade Testing Program, 1968). A table of random numbers was used to implement this plan. Group size was distributed uniformly over the different effectiveness classifications. The total group consisted of 9087 individuals, and group size ranged from 20 to 399, with an average group size of 168.28.

Finally, individual student scores were generated randomly within each group. After the groups were formed, the sample mean for each group was calculated. Some reranking of the groups occurred as a result of these sample values. Sixteen groups were designated as "effective", 20 as "average", and 18 as "less effective". These sample values are the ones which would be observed directly in actual settings. The general characteristics of the sample data are given in Table 3. The average within-group standard deviation was 86.92.

Insert Table 3 about here

Results

A regression equation was obtained using the sample input mean and the sample output mean for each group as the predictor and criterion, respectively. The standard error of estimate was 27.07, and the predictor sum of squares was 57230.16.

Insert Table 4 about here

The distribution of Performance Indicators established for each of the 54 groups using Equation 1 and Equation 2 is given in Table 4. Fifteen of the 16 "effective" groups received a PI of 5, and all of the 18 "less effective" groups received a PI of 1. Groups having observed outputs greater than 10.05 units or less than -10.05 units from the regression line received PIs of 5 and 1, respectively. Groups having observed outputs which deviated between 3.35 and 10.05 units from the line received PIs of 4 and 2.

Insert Table 5 about here

Table 5 shows the number of groups having observed outputs above, within, and below the hyperbolic confidence bands constructed on the regression line using Equation 3 with confidence levels no less than .75, .90, .95, and .99, respectively. Generally, groups having a PI of 1 or a PI of 5 fell outside the bands for each confidence level, and

groups having a PI of 2, 3, or 4 fell inside the bands. When the confidence was increased to .95, all points outside the band corresponded to groups with a PI of 1 or 5. At confidence levels of .75 and .90, six and two points with PIs different than 1 and 5, respectively, fell outside the band. Generally, the difference between the confidence bands of .90, .95, and .99 is slight in terms of location of groups.

Insert Table 6 about here

Comparisons between the absolute distances from the regression line to the Scheffe and Gafarian confidence bands are shown in Table 6. The Gafarian bands were calculated using an extended lower input value so that the input mean would be at the midpoint of the interval over which the bands were constructed. If the actual observed extreme values were to be used, the absolute distance from the regression line to these bands would decrease by about two units. However, this latter condition would violate a condition on which the tables provided by Gafarian (1964) are based.

For most of the interval of interest, the Gafarian bands are wider than the Scheffe bands for each confidence level. The relationship between the Gafarian bands and PIs

is given in Table 7. From Table 5 and Table 7, it appears that the Gafarian technique would result in more conservative decisions about differences in relative effectiveness than would the Scheffé technique.

Insert Table 7 about here

Discussion

When viewed in the light of the Scheffé bands at the usual confidence levels, the observed outputs of groups with a PI of 2, 3, or 4 could not be considered different from the predicted values. However, most of the groups with a PI of 1 or 5 did lie outside of the confidence bands. Thus, these extreme values of PIs could be used to identify schools which are achieving below and above expectation. It appears that attempts to make finer discriminations using PIs may not be warranted. Perhaps a classification system using the categories: 1) achieving about predicted; 2) achieving around predicted; and 3) achieving below predicted may be more realistic than a five category classification like the PIs.

For these data, the hyperbolic Scheffé confidence bands were narrower than the linear confidence bands of Gafarian for most of the interval of interest at each confidence level. Since the width of a confidence band is related directly to errors of prediction, narrower bands are

preferred to wider ones at a given confidence level. In many practical situations, the interval of interest is probably such that the Scheffé bands will be more appropriate than the Gafarian bands. Exceptions may occur if interest centers on schools whose input scores are in a small neighborhood about the input mean. Gafarian (1964) indicates those circumstances in which the linear bands will be more efficient than the Scheffé bands. The efficiency criterion used by Gafarian is the minimization of the area encompassed by the bands.

Overall, PIs and Scheffé bands seem to yield comparable results. It is difficult on the basis of the data presented here to consider either method better in a cost-effective sense. PIs do require use of the average within-school standard deviation on the output variable and the average school size. In some instances, the former information may not be readily available. Forsyth (1973) suggests a method to estimate the former in the event that complete within-school statistics are not available.

The results of this study should assist investigators in developing and implementing a strategy for determining the relative effectiveness of schools. How different techniques might be used to determine relative effectiveness

is dictated somewhat by the intent of the investigator. A liberal strategy would be appropriate if one wished to identify all possible pairs of schools which differed in effectiveness, even at the risk of designating as different some which really did not differ in effectiveness. If false designations are considered serious, then a more conservative strategy would be appropriate. For the Scheffé procedure, a more conservative strategy would dictate the choice of higher confidence levels. For PIs, a conservative strategy would dictate adopting a decision rule which requires a difference of 4 PI units between two schools in order to consider them different in effectiveness. Use of a less stringent decision rule would constitute a more liberal strategy.

Since the data used in this study were generated artificially, the results should be interpreted with some caution. The methods reviewed in this paper need to be applied to longitudinal data from real settings to determine if the results are comparable to those found in this study.

References

Burke, H. R. A study in public school accountability through the application of multiple regression to selected variables. (Doctoral dissertation, Indiana University, 1972). Dissertation Abstracts International, 1973, 34, 4661A-4662A. (University Microfilms No. 73-6965)

California Test Bureau/McGraw-Hill. Technical report: Comprehensive Tests of Basic Skills. Monterey: McGraw-Hill, 1970.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., and York, R. Equality of educational opportunity. Washington, D.C.: U. S. Government Printing Office, 1966.

Convey, J. J. A validation of three models for producing school effectiveness indices. A paper presented to the annual meeting of the American Educational Research Association, Washington, D.C., 1975. (Eric Document Reproduction Service No. ED 106 368)

Dyer, H. The Pennsylvania Plan. Science Education, 1966, 50, 242-248.

Dyer, H., Linn, R., & Patton, M. Feasibility study of educational performance indicators: Final report to New York State Education Department. Princeton: Educational Testing Service, 1967.

Dyer, H., Linn, R., & Patton, M. A comparison of four methods of obtaining discrepancy measures based on observed and predicted school system means on achievement tests. American Educational Research Journal, 1969, 4, 591-605.

Florida Ninth-Grade Testing Program. Technical report: 6-68. Tallahassee: Florida State University, 1968.

Forsyth, R. Some empirical results related to the stability of performance indicators in Dyer's Student Change Model of an educational system. Journal of Educational Measurement, 1973, 10, 7-12.

Gafarian, A. V. Confidence bands in straight line regression. Journal of the American Statistical Association, 1964, 59, 182-213.

Guthrie, J. W. A survey of school effectiveness studies. In U. S. Department of Health, Education, and Welfare, Do teachers make a difference? Washington: U. S. Government Printing Office, 1970.

Lindquist, E. F. Factors determining reliability of test norms. Journal of Educational Psychology, 1930, 21, 512-520.

Lord, F. M. Test norms and sampling theory. Journal of Experimental Education, 1959, 27, 247-263.

Marco, G. L. A comparison of selected school effectiveness measures based on longitudinal data. Journal of Educational Measurement, 1974, 11, 225-234.

Maryland State Department of Education. Maryland accountability program report. Baltimore: Maryland State Department of Education, 1975.

Miller, R. G. Simultaneous statistical inference. New York: McGraw-Hill, 1966.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Shaycoft, M. F. The high school years: Growth in cognitive skills. Pittsburgh: American Institute of Research, 1967.

Working, H. & Hotelling, H. Applications of the theory of error to the interpretation of trends. Journal of the American Statistical Association, 1929, 24, 73-85.

Table 1

Specifications for Initial 54 Ordered Sets

Variable	Mean	S.D.	Correlation
Output	539	34.36	.73
Input	474	27.32	

Table 2

Characteristics of Ordered Sets After Pairing

Variable	Mean	S.D.	Correlation
Output	535.37	40.06	.7712
Input	477.70	31.91	

Table 3

Characteristics of Sample Data

Variable	Mean	S.D.	Correlation
Output	535.82	41.99	.7695
Input	477.43	32.86	

Table 4

Distribution of PIs Over Effectiveness

Classifications Based on Sample Data

PI	Effective	Average	Less Effective
5	15	0	0
4	1	8	0
3	0	7	0
2	0	4	0
1	0	1	18

Table 5

Distribution of PIs Relative to Scheffe Confidence Bands

PI	.75 Confidence			.90 Confidence			.95 Confidence			.99 Confidence		
	Above	Within	Below	Above	Within	Below	Above	Within	Below	Above	Within	Below
5	15	0	0	15	0	0	14	1	0	13	2	0
4	4	5	0	1	8	0	0	9	0	0	9	0
3	0	7	0	0	7	0	0	7	0	0	7	0
2	0	2	2	0	3	1	0	4	0	0	4	0
1	0	0	19	0	1	18	0	1	18	0	1	18
	19	14	21	16	19	19	14	22	18	13	23	18

Table 6

Distance From Regression Line To

Scheffé and Gafarian Bands

Confidence	Scheffe			Gafarian
	Upper Extreme	Mean	Lower Extreme	
.75	16.48	6.23	14.15	13.94
.90	21.43	8.10	18.14	18.68
.95	24.61	9.31	21.13	22.20
.99	31.06	11.74	26.66	28.43

Table 7

Distribution of PIs Relative to Gafarian Confidence Bands

PI	.75 Confidence			.90 Confidence			.95 Confidence			.99 Confidence		
	Above	Within	Below	Above	Within	Below	Above	Within	Below	Above	Within	Below
5	14	1	0	11	4	0	10	5	0	9	6	0
4	0	9	0	0	9	0	0	9	0	0	9	0
3	0	7	0	0	7	0	0	7	0	0	7	0
2	0	4	0	0	4	0	0	4	0	0	4	0
1	0	1	18	0	2	17	0	5	14	0	11	8
	14	22	18	11	26	17	10	30	14	9	37	8